ACVM - 5. Assignment - Long-Term Tracking

Nejc Ločičnik, 63180183, nl4952@student.uni-lj.si

I. Introduction

In this assignment, we upgrade a short-term tracker SiamFC into a long-term one. We discuss how to evaluate tracking confidence, different types of sampling methods and amount of sampled regions, all of these influence how the re-detection part of a long-term tracker operates. The tracker is evaluated on the sequences provided with the assignment. Taking into consideration the computationally intensive CNN-based tracker SiamFC and since I don't have access to a Cuda GPU, the testing was only done on the *car9* sequence.

II. Experiments

A. Short-term tracker SiamFC

There isn't much to say about the short-term SiamFC tracker. This section is merely a confirmation that all of the provided material was setup correctly. The results obtained (table I) are similar to the ones on the assignment slides.

Sequence	Precision	Recall	F-Score
car9	0.636	0.270	0.379

Table I
Results of the short-term variation of SiamFC.

B. Long-term tracker SiamFC

To upgrade the short-term tracker into a long-term tracker we need to introduce a measure of confidence, so we can decide when exactly is re-detection necessary and the re-detection mechanism itself. The re-detection mechanism is done by randomly (uniform or normal distribution) sampling regions either on the whole picture or localised around the last confident position.

C. Confidence score

SiamFC localizes target by correlation, so we can simply take the max/peak value of the response. The value itself can be interpreted as the likelihood that the target is there, so we only need to determine where we threshold it. We could have two threshold to determine full occlusions or partial occlusions. For example in figure 1, which plots peak values in a sequence, we can clearly see that there must be a full occlusion at frames 750-850. The downturn at the end is because the predicted bounding box doesn't reshape fully compared to the tracking object deformation (camera perspective change), it does resize though. The optimal value that produced the best results was 5 (also shown on the plot).

D. Region sampling

I tried uniform sampling and sampling using a normal (Gaussian) distribution. For the uniform sampling I used a variation using Hilton sequence, which produces a corrected (more uniform) distribution of samples compared to the standard uniform sampling (the samples are more evenly spaced out). An example of sampling is shown in figure 2.

I didn't even attempt sampling over the whole image. I feel like to get a reasonable coverage of the image you need to heavily increase the amount of samples, which increases required computation. Instead I localized the sampling to the last confidently tracked position (direct sampled regions towards an approximate location of the target). The sampling area is initially 200 pixels in each direction. This area slowly increases as long as the re-detection process lasts (+1 or 2 for a failed re-detection). The "area" for Gaussian sampling of a bit different, for that we set: σ = last confident location, $\mu = 200$ (μ grows).

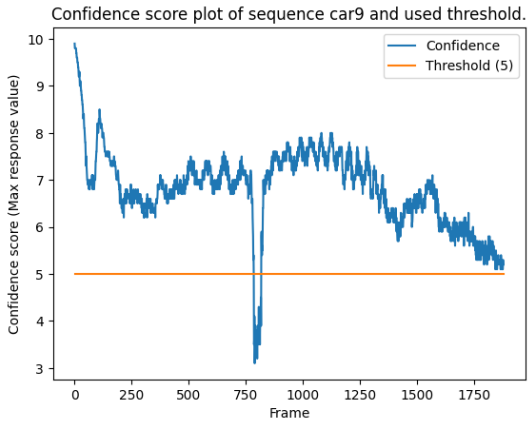

Figure 1. Plot of confidence (peak response value) for sequence car9 and threshold (5).

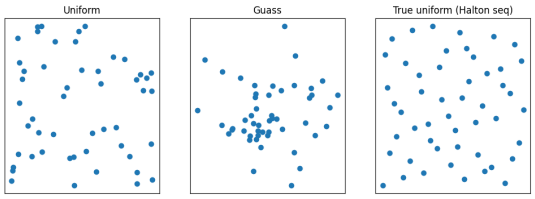

Figure 2. Comparison of different sampling methods (N=50).

An example of the whole re-detection process is shown on figure 3. Bold green square shows the ground truth, the bold red square shows the predicted tracking position and the yellow squares show the sampled regions. This example runs Gaussian sampling with 30 samples (as seen Gaussian samples often overlap, which lowers computational efficiency). In frame 785 the tracking confidence drops slightly below the threshold, which means re-detection is enabled. Frames 800 and 810 showcase region sampling during re-detection (when the tracking object is occluded). Frame 818 demonstrates successful re-detection.

E. Number of sampled regions

Increasing the number of sampled regions does different things whether we use the corrected uniform distribution (Hilton sequence) and normal (Gaussian) distribution. For the uniform sampling, increasing the number of samples gives us a better representation of the sampled area (since the samples are evenly distributed). While for the Gaussian sampling we get a higher guarantee of sampling regions surrounding the area of interest (e.g. last confident position). For high amount of samples with Gaussian sampling we also have an increased

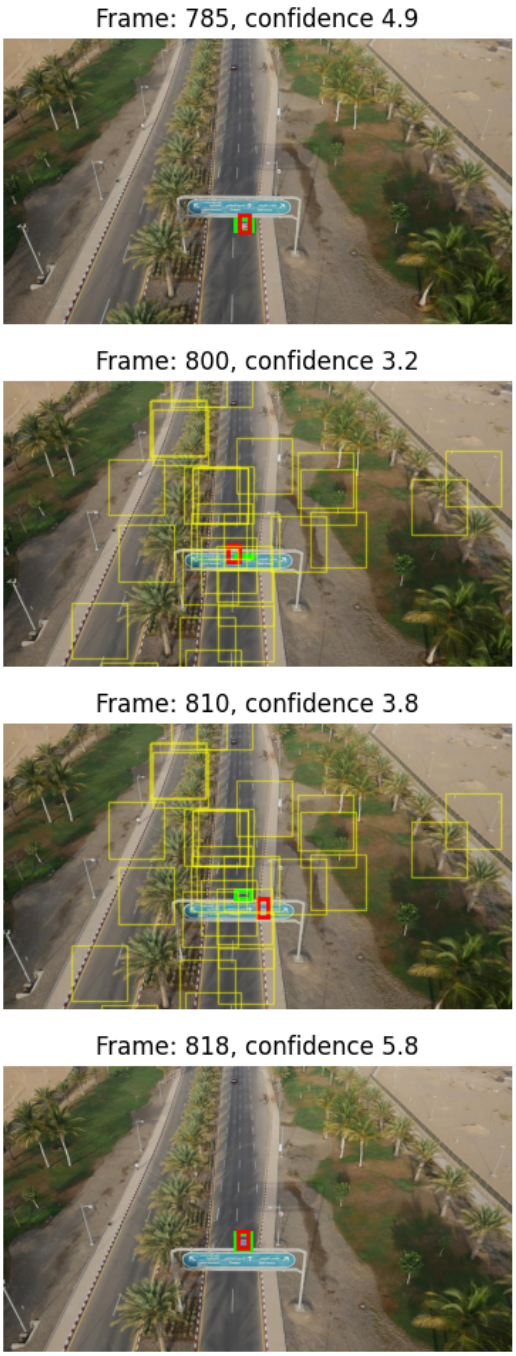

Figure 3. An example of re-detection on sequence car9 (Gaussian sampling, samples=30, std=200).

chance of overlaps, meaning redundant computation (this can be avoided with some post-processing after sampling, but I choose not to do this).

So in my opinion, when we fully lose the target and want to sample the whole image, the corrected uniform sampling with the correct amount of samples (enough for image representation) should produce better results. When we know the approximate location of the target (like the last confident position), then Gaussian sampling should produce better results.

Table II-E shows results for different amount of samples using the corrected uniform distribution and the normal distribution. As expected, a higher amount of sampled regions **increases the chance** that the target will be re-detected, boosting recall and f-score. But realistically as seen in the table, the difference is actually really small as long as we have enough samples to reliably sample the required area.

Sampler	N	Area	Precision	Recall	F-Score
Gauss	30	150	0.603	0.591	0.597
Gauss	20	150	0.602	0.590	0.596
Gauss	10	150	0.602	0.578	0.590
Corr. Uni.	30	200	0.600	0.588	0.594
Corr. Uni.	20	200	0.598	0.586	0.592
Corr. Uni.	10	200	0.600	0.589	0.595

Table II
RESULTS FOR VARIED AMOUNT OF SAMPLED REGIONS (GAUSSIAN AND CORRECTED UNIFORMED SAMPLING).

III. CONCLUSION

In this assignment, we successfully extended the SiamFC short-term tracker to a long-term tracker by introducing a confidence measure and re-detection mechanism. By evaluating different region sampling methods, such as uniform and Gaussian distributions, we found that both methods have their advantages depending on the scenario. Corrected uniform sampling is more effective for full image coverage, while Gaussian sampling is advantageous when the approximate target location is known. Our experiments on the *car9* sequence demonstrated the effectiveness of these sampling methods and highlighted the importance of choosing an appropriate number of samples to balance computational efficiency and re-detection accuracy. Overall, the enhanced SiamFC tracker shows improved performance in long-term tracking tasks.